An Entropic View of Pickands' Theorem

Jean-François Bercher
Laboratoire des Signaux et Systèmes,
CNRS-Univ Paris Sud-Supelec,
91192 Gif-sur-Yvette cedex, France
Email: bercherj@esiee.fr

Christophe Vignat
Institut Gaspard Monge
Université Paris-Est
77454 Marne-la-Vallée cedex 02, France
Email: vignat@univ-mlv.fr

ABSTRACT

It is shown that distributions arising in Rényi-Tsallis maximum entropy setting are related to the Generalized Pareto Distributions (GPD) that are widely used for modeling the tails of distributions. The relevance of such modelization, as well as the ubiquity of GPD in practical situations follows from Balkema-De Haan-Pickands theorem on the distribution of excesses (over a high threshold). We provide an entropic view of this result, by showing that the distribution of a suitably normalized excess variable converges to the solution of a maximum Tsallis entropy, which is the GPD. This highlights the relevance of the so-called 'Tsallis' distributions in many applications as well as some relevance to the use of the corresponding entropy.

I. INTRODUCTION

Generalized Pareto Distributions (GPD) are widely used in practice for modeling the tails of distributions. The underlying rationale is the Balkema-De Haan-Pickands theorem [**?**], [**?**], which asserts that the distribution function of the excess variable $X-u|X>u$ (that is the distribution of the (shifted) variable X exceeding a threshold u) converges to a GPD with survival function, as $u \rightsquigarrow \infty$:

$$S_X(x) = Pr\left(X>x\right) = \left(1+\frac{\gamma}{\sigma}x\right)^{-\frac{1}{\gamma}}, \qquad (1)$$

where σ is a scale parameter and γ a shape parameter. For $\gamma=0$, the GPD reduces to the exponential distribution.

In applied fields, GPD have encountered a large success since they were obtained as the maximizers of a special entropy, the Tsallis (Harva-Charvat-Daróvczy) entropy [**?**], with suitable constraints. It is worth mentioning that monotonous transforms of the latter entropy also exhibit the same GPD maximizers: an important example is Rényi entropy [**?**]. Of course, in a wide variety of fields, experiments, numerical results and analytical derivations fairly agree with the description by distribution (1). This distribution is of very high interest in many physical systems, since it enables to model power-law phenomena. Indeed, power-laws are especially interesting since they appear widely in physics, biology, economy, and many other fields [**?**].

In this communication, we give an interpretation of Pickands' result which links it to the maximum (Rényi/Tsallis) entropy setting with emphasis on the the importance of GPD for modeling excess distributions. This view gives a possible interpretation for the ubiquity of 'Tsallis' (GPD) distributions in physics applications, as well as in other fields, an an argument in support to the use of Rényi/Tsallis entropies.

In the following, we deal with univariate distributions defined on $\mathbb{R}$ or on a subset of $\mathbb{R}$. Our approach is as follows. First, we show that the GPD can be obtained as the solution of as resulting of a Rényi-Tsallis maximum entropy problem, with normalization and moment constraints. Second, we consider distributions in the domain of attraction of Fréchet distributions. This family includes for instance Cauchy, Student and Pareto distributions. We characterize the associated q-norm and first moment of the survival function associated to the excess variable $X-u|X>u$. Using an appropriate normalization, we define a variable whose survival function q-norm and moment converge to constant values. Third, we show that it is possible to exhibit a parameter q, as a function of the the parameter of the Fréchet family, such that the maximum entropy and values of the constraints coincide with those of the normalized variable from the Fréchet family. Therefore, since the maximum entropy with given constraints is unique, we obtain that the excess variable from a distribution in the domain of attraction of Fréchet distribution asymptotically follows a Generalized Pareto Distribution (1).

II. SOLUTION TO THE MAXIMIZATION OF TSALLIS' ENTROPY

We first begin by the expression of the solution to the maximization of Tsallis' entropy subject to normalization and moment constraints.

Theorem 1: Consider the set $\mathcal{F}=\{G:R^+\rightarrow\mathbb{R}\}$

The maximum Tsallis entropy problem (or equivalently the maximum q-norm problem), with $q<1$, defined by

$$\max_{G\in\mathcal{F}} S_q(G) = \max_{G\in\mathcal{F}} \frac{1}{1-q}\left(\int_{\mathcal{D}} G(z)^q dz - 1\right)$$

subject to

$$\int_0^{+\infty} zG(z)dz = \mu \text{ and } \int_0^{+\infty} \mathrm{G(z)dz} = \theta \qquad (2)$$

has for unique solution

$$G_*(z) = \alpha^{\frac{1}{q-1}} \left(1 + \frac{\beta}{\alpha} z\right)^{\frac{1}{q-1}} \tag{3}$$

with, for $q \geq 1/2$

$$\mu = \frac{(q-1)^2}{q(2q-1)} \frac{\alpha^{\frac{2q-1}{q-1}}}{\beta^2},\ \theta = \frac{\alpha^{\frac{q}{q-1}}}{\beta} \frac{(1-q)}{q} \tag{4}$$

$$\text{and } S_q(G_*) = \frac{\alpha^{\frac{2q-1}{q-1}}}{\beta} \frac{(1-q)}{(2q-1)}. \tag{5}$$

The mean is not defined for $q < 1/2$.

Proof: We follow here the approach of [**?**]. Consider the functional Bregman divergence:

$$B(f,g) = \int d(f,g) dx \tag{6}$$

$$= \int -\left(f(x)^q - g(x)^q - q\left(f(x) - g(x)\right) g(x)^{q-1}\right) dx \tag{7}$$

associated to the (pointwise) Bregman divergence $d(f,g)$ built upon the strictly convex function $-x^q$ for $q \in (0,1)$. Then let us evaluate the divergence between the distribution $G_*(z)$ in (3) and any distribution $G(z)$, with G dominated by G_*, $G(z) \ll G_*(z)$, and satisfying (2):

$$B(G, G_*) = -\int_{\mathcal{S}} G(z)^q - G_*(z)^q - \alpha(G(z)G_*(z)^{q-1} - G_*(z)^q) dz \tag{8}$$

$$= -\int_{\mathcal{S}} G(z)^q dz + \int_{\mathcal{S}} G_*(z)^q dz, \tag{9}$$

where $\mathcal{S}$ denotes the support of $G_*(z)$. The second line follows from the fact that since G and G_* both satisfy (2), then, using (3) it is easy to check that

$$\int_{\mathcal{S}} G(x) G_*(x)^{q-1} dx = \int_{\mathcal{S}} G_*(x)^q dx.$$

The Bregman divergence $B(G, G_*)$ being always positive and equal to zero if and only if $G = G_*$, the equality (9) implies that, for $q \in (0,1)$,

$$S_q(G_*) \geq S_q(G) \tag{10}$$

which means that G_* is the distribution with maximum Rényi-Tsallis entropy, with $q \in (0,1)$, in the set of all distributions $G \ll G_*$ satisfying the constraints (2). Values of the constraints (4) and of the maximum entropy (5) follow by direct calculation. ■

III. The distribution of excesses for distributions in Fréchet family

In the following, we consider the set of distributions which belongs to the Fréchet domain of attraction. This is the set of all distributions $\mathcal{F}$ such that if variables X_i are independent and identically distributed according to a distribution in $\mathcal{F}$, then $\max_{i=1..N} X_i$ converges to the Fréchet distribution. This family typically represents heavy-tailed distributions whose tail behave as a power-law. It was shown by Gnedenko [**?**] that a necessary and sufficient condition for a distribution to be in the Fréchet domain of attraction is that its survival function satisfies

$$\lim_{z \to +\infty} \frac{S(z)}{S(cz)} = c^a,$$

with $c > 0$, $a > 0$. Equivalently, this can also read

$$\lim_{z \to +\infty} S(z) = z^{-a} l(z),$$

where $l(z)$ is a slowly varying function, i.e. a function such that $\lim_{z\to+\infty} \frac{l(zt)}{l(z)} = 1,\ \forall t > 0$.

Let us consider the excess variable $X_u = X - u | X > u$. Its survival function is

$$S_{X_u}(z) = \frac{S_X(z+u)}{S_X(u)}.$$

Theorem 2: Suppose that X_u belongs to the Fréchet domain, with

$$S_{X_u}(z) \sim z^{-a} l(z),$$

then its $q-$norm is asymptotically

$$\|S_{X_u}\|_q \sim \frac{u}{aq-1}$$

and its first moment is asymptotically, with $a \geq 2$,

$$\int_0^{+\infty} z S_{X_u}(z) dz = \frac{u^2}{(1-a)(2-a)}.$$

Proof: the $q-$norm writes

$$\begin{aligned}
\|S_{X_u}\|_q &= \int_0^{+\infty} \left(\frac{S_X(z+u)}{S_X(u)}\right)^q dz \\
&= \int_u^{+\infty} \left(\frac{S_X(z)}{S_X(u)}\right)^q dz \\
&= u \int_1^{+\infty} \left(\frac{S_X(wu)}{S_X(u)}\right)^q dw \\
&\sim u \int_1^{+\infty} \frac{(uw)^{-aq}}{u^{-a}} \left(\frac{l(wu)}{l(u)}\right)^q dw \\
&\sim u \int_1^{+\infty} \frac{(uw)^{-aq}}{u^{-a}} dw \\
&= u \int_1^{+\infty} w^{-aq} dw = \frac{u}{aq-1},
\end{aligned}$$

with $1 - aq \leq 0$, since $a \geq 2, q \geq 1/2$. Of course, we immediately obtain, taking $q = 1$, that

$$\|S_{X_u}\|_1 = \frac{u}{a-1}.$$

Similarly, the first moment is

$$\begin{aligned}
\int_0^{+\infty} z S_{X_u}(z)\,dz &= \int_0^{+\infty} z\frac{S_X(z+u)}{S_X(u)}dz \\
&= \int_u^{+\infty} (z-u)\frac{S_X(z)}{S_X(u)}dz \\
&= \int_1^{+\infty} u(w-1)\frac{S_X(wu)}{S_X(u)}udw \\
&\sim u^2\int_1^{+\infty}(w-1)\,w^{-a}dw \\
&= \frac{u^2}{(1-a)(2-a)}.
\end{aligned}$$

■

We have a simple corollary to this theorem:

Corollary 1: The survival function S_{Y_u} of random variable $Y = X/g(u)$, where function g is such that $g(u) \sim u$, has asymptotical norms

$$\|S_{Y_u}\|_q \sim \frac{1}{aq-1} \text{ and } \|\mathrm{S}_{\mathrm{Y_u}}\|_1 = \frac{1}{\mathrm{a}-1}.$$

and an asymptotical first moment

$$\int_0^{+\infty} z S_{Y_u}(z)\,dz \sim \frac{1}{(1-a)(2-a)}.$$

Proof: The results for S_{Y_u} follow directly from Theorem 2, with

$$S_{Y_u}(z) = S_{X_u}(zg(u)). \tag{11}$$

■

IV. The entropy of the distribution of excesses

Of course, coincidence of the survival function $S_{Y_u}(z)$ of the normalized excess Y_u with the maximum entropy solution (3) imposes

$$a = \frac{1}{1-q}.$$

Then, it is easy to check that

Theorem 3: if X belongs to the Fréchet family, then choosing $q < 1$ such that

$$a = \frac{1}{1-q},$$

the excess distribution of Y_u reaches asymptotically the maximum $q-$norm solution under constraints asymptotically equal to μ and θ provided $\alpha = \beta = 1$.

Proof: Choosing $a = \frac{1}{1-q}$ yields

$$\|S_{Y_u}\|_q \sim \frac{1}{aq-1} = \frac{1-q}{2q-1},$$

$$\|S_{Y_u}\|_1 \sim \frac{1}{a-1} = \frac{1-q}{q}$$

and

$$\int_0^{+\infty} z S_{Y_u}(z)\,dz \sim \frac{1}{(1-a)(2-a)} = \frac{(q-1)^2}{q(2q-1)}$$

which coincide with the unique maximum $q-$norm function with constraints μ and θ if and only if $\alpha = \beta = 1$. ■

We can now return to the distribution of the excess without normalization. Then we obtain a Generalized Pareto Distribution, whose shape parameter is given by the exponent parameter of the Fréchet family, and where the scale parameter is a function of the threshold u and of the Fréchet parameter a. This can be stated as follows.

Theorem 4: If X belongs to the Fréchet family, then the survival function of its excesses over a threshold u converges (in the $q-$norm sense) to

$$G(z) = \left(1 + \frac{\gamma}{\sigma}x\right)^{-\frac{1}{\gamma}} \tag{12}$$

with $\gamma = 1/a$ and $\sigma = g(u)/a$ for any positive function g such that $g(u) \sim u$.

Proof: The proof is immediate by (11). ■

At this point it is still important to emphasize that the GPD (12) enjoys a threshold stability property: the distribution of the excesses over a threshold of GPD remains a GPD, with the same exponent but a different shape parameter. This stability result is certainly also a reason of the ubiquity of GPD in many applications.

Theorem 5: Given a GPD with parameters γ, σ the distribution of excesses remains a GPD with parameters γ and $\sigma' = \sigma\left(1 + \frac{\gamma}{\sigma}u\right)$.

Proof: As usual, let u denotes the threshold, S_X the survival function of the original GPD and $S_{X_u}$the survival function of variable X_u.Then,

$$S_{X_u}(x) = \frac{S_X(x+u)}{S_X(u)} = \frac{\left(1+\frac{\gamma}{\sigma}(x+u)\right)^{-\frac{1}{\gamma}}}{\left(1+\frac{\gamma}{\sigma}u\right)^{-\frac{1}{\gamma}}} = \left(1+\frac{\gamma}{\sigma'}x\right)^{-\frac{1}{\gamma}}$$

where the last expression is obtained after having factored the term $\left(1+\frac{\gamma}{\sigma}u\right)^{-\frac{1}{\gamma}}$in the numerator.

■

We have shown that the distribution of excesses in the Fréchet family, which is a Generalized Pareto Distribution, can be interpreted as a maximum q-entropy (or maximum q-norm) solution. With this result, it is possible to connect the ubiquity of heavy-tailed distributions in physics, economics or signal processing, the distribution of the excesses over a (often unknown) threshold, and a maximum entropy construction. The final paper will present a numerical illustration in the Student case, and will include a simple extension to the Weibull family, wich leads to distributions with exponential tails.

References

[1] J. Pickands, "Statistical inference using extreme order statistics," *The Annals of Statistics*, vol. 3, pp. 119–131, 1975.

[2] A. A. Balkema and L. de Haan, "Residual life time at great age," *The Annals of Probability*, vol. 2, pp. 792–804, Oct. 1974.

[3] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *Journal of Statistical Physics*, vol. 52, pp. 479–487, July 1988.

[4] A. Rényi, "On measures of entropy and information." Berkeley, Calif.: Univ. California Press, 1961, pp. 547–561.

[5] M. E. J. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics*, vol. 46, pp. 323–351, 2005.

[6] C. Vignat, A. Hero, and J. A. Costa, "About closedness by convolution of the Tsallis maximizers," *Physica A*, vol. 340, pp. 147–152, Sept. 2004.

[7] B. V. Gnedenko, "Sur la distribution limite du terme maximum d'une série aleatoire," *Annals of Mathematics*, vol. 44, pp. 423–453, 1943.